

創価大学  
国際仏教学高等研究所  
年 報

平成29年度  
(第21号)

Annual Report  
of  
The International Research Institute for Advanced Buddhology  
at Soka University

for the Academic Year 2017

Volume XXI

創価大学・国際仏教学高等研究所  
東京・2018・八王子

The International Research Institute for Advanced Buddhology  
Soka University  
Tokyo・2018

## From philology to history: Deciphering the language of ancient Afghanistan\*

Nicholas SIMS-WILLIAMS

One of the books which first roused my interest in pursuing the study of ancient languages was John Chadwick's "The decipherment of Linear B" — a wonderful tale, as exciting as a detective story, but with the additional advantage of describing the solution to a real-life mystery rather than one invented by the author. The story of Michael Ventris's decipherment of Linear B is a story of the most difficult type of decipherment, involving a completely unknown script and a language which was also at the time unknown (though of course it eventually turned out to be an early form of Greek). I cannot promise you that the story I have to tell you today, that of the rediscovery of the ancient language of Afghanistan, will be equally exciting, but there are many parallels. The decipherment of the Linear B tablets not only revealed a form of the Greek language far older than any known before, but also cast new light on the earliest Greek poetry and the history of Greece; similarly, the decipherment of Bactrian, as we now call it, has given us a previously unknown language and has begun to fill in the gaps in our very imperfect knowledge of the ancient history and culture of Afghanistan and adjacent lands. By telling you this story, I hope to demonstrate what philology can achieve: in particular, how a text which is at first completely incomprehensible can be made to give up its secrets by patient, systematic analysis. But I must admit straight-away that the decipherment of Bactrian was not nearly so difficult as the decipherment of Linear B: although the Bactrian language was indeed unknown, it is written in a script which was already at least partially known, a local variety of the Greek alphabet. I should really say: *two* local varieties of the Greek alphabet, since it appears in two substantially different forms, one "monumental" and one "cursive". So there are really two stories to tell: the first about the discovery and interpretation of the Bactrian inscriptions in monumental script, the second about the later decipherment of the cursive script.

One of the earliest records of Bactrian, an inscription of the 2nd century AD, refers to the language as *ariao*, that is, "Aryan", a term which we can hardly use nowadays—not only because of its political overtones, but also because it is equally applicable to any language of the Iranian family: Darius the Great had used the same name to refer to the language which we now call Old Persian. Later, in early Islamic times, by which time Bactria was renamed Tukharistan, the language was known as "Tukhari" or "Tocharian", but modern scholarship

---

\* Lecture delivered at IRIAB, Soka University on 21st October 2017. It was previously given at the Norwegian Academy in Oslo on 7 March 2017 and published in the Academy's Yearbook for 2017.

has appropriated that name for a completely different group of Indo-European languages. So today the language of ancient Afghanistan is universally known as “Bactrian”. As the name implies, the language is assumed to be that of ancient Bactria, the land which lies between the River Oxus or Amu Darya and the Hindukush mountains of central Afghanistan, with its capital at Balkh, a city known to the ancient Greeks as Bactra. The great majority of the Bactrian manuscripts and inscriptions which we know today derive from this very area (see Map).

The Bactrian language belongs to the Iranian branch of the Indo-European family, being fairly closely related to Persian, Pashto and many other languages spoken in Afghanistan today, more distantly to Sanskrit, and of course ultimately to English and most other languages of Europe. Amongst the languages of the Middle Iranian period, that is, approximately the first millennium AD, Bactrian occupies an intermediate position between the Western group, that is, Middle Persian and Parthian, and the Eastern group, consisting of Sogdian, Choresmian, Khotanese and Tumshuqese. Naturally enough, it has most in common with its nearest neighbours, Sogdian and Parthian.

Like most of the older Iranian languages, both Sogdian and Parthian are written in scripts derived from Aramaic. Bactrian, however, is written in Greek script, a legacy of the conquest of Bactria by Alexander of Macedon in the 4th century BC. The successors of Alexander introduced Greek as the language of their administration, and in recent years a number of Greek administrative documents have been found in Afghanistan. After the collapse of Greek rule in Bactria, the first centuries AD saw the growth of the Kushan empire under kings such as Kanishka I, who ruled much of northern India and Central Asia from his powerbase in Bactria, and who was the first to use Bactrian in place of Greek on his coins. In the 3rd century, Bactria was conquered by the Sasanian dynasty of Iran, then by various nomadic peoples including Huns and Turks, before eventually falling to the armies of Islam in the 7th-8th centuries. Bactrian was in use as a written language up to this time, and even a little later, so its recorded history lasts for about 800 years.

We may begin the story of the rediscovery of Bactrian towards the end of the 19th century. At that time, not a single substantial Bactrian text had yet come to light. In so far as the language was known at all, it was from short legends on coins and seals, in particular those of the Kushan period, the 1st to 3rd centuries AD, written in what we now refer to as the “monumental” script. For a scholar with a classical education — and a hundred years ago that would have been every scholar in Europe — the script is quite easy to read. On the other hand, these short inscriptions don’t tell us much about the Bactrian language. They contain names and titles of kings and deities, but virtually no inflected forms and no verbal forms at all; hardly anything, in fact, to give us an idea of Bactrian morphology or syntax.

The status of Bactrian as an unknown language began to change almost sixty years ago, on the 6th May 1957, with the discovery of the first substantial Bactrian inscription at the site of Surkh Kotal. The inscription is 25 lines long, neatly written and perfectly preserved. But although it was easily legible, there were two major problems: the text was written continuously, with no gaps between the words; and almost all of those words were of unknown meaning. The publication was entrusted to a young Belgian scholar, André Maricq, who made the text available almost immediately, in 1958, providing an almost perfect reading of the letters and making a good stab at dividing the text into words; but he didn’t get far with translating it.

Soon afterwards, in 1960, two scholars independently, but more or less simultaneously, published new interpretations of the whole inscription. The first was Helmut Humbach, something of an *enfant terrible*, who had already made a name for himself for his iconoclastic reinterpretation of the most ancient work of Iranian literature, the Gathas of Zarathushtra. According to Humbach, the inscription is a Mithraic hymn, in eight strophes of three to four lines each, in which king Kanishka is simultaneously identified as the son of Mithra and as the god Mithra himself. The second was W. B. Henning, perhaps the greatest specialist in the Middle Iranian languages, according to whom the inscription deals with the foundation of a temple by Kanishka, its abandonment because of problems with the water supply, the digging of a well and the re-establishment of the temple by an official named Nokonzoko.

Everything we have since learned about Bactrian confirms that Henning's more down-to-earth version was essentially correct. But how could two scholars come to such radically different results? They had the same text in front of them, and both shared the same assumption that the text was written in the Middle Iranian language of Bactria, at that time effectively unknown. The same methods were open to both of them: context, etymology and the rules of historical phonology.

As an example of Henning's use of these methods I would like to quote two short passages as he translated them. (You will see that where he had nothing plausible to suggest he prudently left some words untranslated.) The first passage describes what happened because of the lack of a water-supply: "... whereby the acropolis came to be waterless ..., then the gods withdrew from the seat ... and the acropolis was abandoned (*pidorigd-o*)". The second describes the intended outcome of Nokonzoko's building works: "... so that through them pure water shall not be lacking to the acropolis ..., may then the gods not withdraw from their seat, and may their acropolis not become abandoned (*pidorixs-ēio*)". In the first passage, as Henning recognized, the verbs are all in the past tense; in the second they are in the present optative. Comparing the two passages, one sees that the two verbal forms with which they end must attest the past and the present stem respectively of one and the same verb. The past stem *pidorigd-* ends with a *d*, the present stem *pidorixs-* with an *s*. The relationship between the two is characteristic of Sogdian and some other Middle Iranian languages, in which past stems end in *d* or *t* (just as in English!) while the suffix *-s* forms intransitive or passive present stems.

Another acute observation of Henning's was that the Greek script had no letter representing a voiceless affricate such as *č* (English *ch*), a very common type of sound in virtually all Iranian languages. As he wrote: "A Middle Iranian language lacking affricates or sounds representing the ancient affricates ... is frankly impossible". Starting from this premise, he recognized that the Old Iranian *č*, however it may have been pronounced in Bactrian, was represented by the Greek letter *sigma*. This made it possible to see that the spelling *sado* could not only represent the word for "100", Old Iranian *\*sata-*, but also the word for "a well", Old Iranian *\*čāt-*. This was a significant result, since the construction of a well turned out to be one of the main topics of the inscription. Henning also recognized this use of the letter *sigma* for older *\*č* in forms such as the preposition *aso* "from" or the relative pronoun *sido* "which" — an equally important result, since it is little words like these which give a text its structure and make it possible to interpret its syntax even if one does not know the meaning of the nouns and verbs.

I will mention just one further expression amongst many for which Henning was the first to find a plausible interpretation: *ōsogdo-maggo*. Maricq had translated “hemp was burnt”, comparing Persian *mang* “hemp” and *soxtan* “to burn”, but Henning recognized that the two words form a compound meaning “pure-minded”, “with a pure heart”, a compound which has a precise cognate in Sogdian. It may have been this very phrase, as understood by Maricq, which set Humbach off in the wrong direction, towards a mystical, religious interpretation of the text. But in any case it seems to me that Humbach’s previous work, which focused on a ritual interpretation of the oldest Iranian and Indian texts, predisposed him to such a viewpoint. Henning’s greater familiarity with the Middle Iranian languages, and the more practical content of most Middle Iranian inscriptions, tended to protect him from such extravagances.

So far I have been talking about the discovery and interpretation of Bactrian coins and inscriptions in the “monumental” script. In this case no real decipherment was required, as the script could already be read. But, as I said at the beginning, there is a second story to be told, about the decipherment of Bactrian texts in cursive script.

Here too, the material that has been known for longest consists of coins and seals, mainly from the time after the Kushan dynasty. At the beginning of last century, when the Kushan coin-legends were already quite well understood, the later legends in cursive script could hardly be read at all: as late as 1901, the *Journal Asiatique* published an attempted decipherment based on the assumption that they were written from right to left, in a variety of Aramaic script, rather than in Greek script from left to right. By 1930 or so, the earliest coin-legends in cursive script could be read fairly correctly, in part because their content—names, titles and so on—was so predictable, but the later coin legends, in a cursive which had developed yet further away from the monumental script, were still largely incomprehensible.

A few scraps of manuscripts on paper written in the latest form of this cursive script had been recovered by German archaeological expeditions to Turfan in western China in the early 1900s, but no-one tried to read them until the 1950s. Unfortunately all of the fragments lack either the right or the left margin, so they don’t contain a single complete line of text between them. That was only one of many problems for the decipherer. Unlike coins, with their largely predictable legends, no assumptions could be made about the content of the manuscripts; and the cursive writing had developed to such an extent that only a few letters could be clearly identified with those of the earlier monumental script.

The first to attempt a reading of these fragments was Olaf Hansen in 1951. With the benefit of hindsight, we can see that he succeeded in correctly identifying ten letters, less than half of the alphabet. Not surprisingly, he did not discover the correct reading of a single word, though he came close in a couple of cases. Some progress was made during the 1960s by Helmut Humbach and by my own teacher, Ilya Gershevitch, himself a student of W. B. Henning. By this time the Surkh Kotal inscription was known, and Humbach and Gershevitch were able to recognize the cursive forms of several words attested there, including basic words such as conjunctions and prepositions. But all in all, the manuscript fragments remained mysterious, and there seemed to be no way of making significant progress.

My own involvement began just a few years after this. From 1968 to 1975 I was Gershevitch’s pupil in Cambridge, studying Sogdian and other Iranian languages, first as an undergraduate and then as a research student. Bactrian was not on the syllabus — in fact I suspect that until I began teaching it in London in the 1990s Bactrian had not been on the

syllabus anywhere for more than a thousand years — but one summer I decided that so little had been written about Bactrian that it would be a manageable task to read it all in the summer vacation. The result was a small discovery about Bactrian syntax, which was published in 1975 in one of my very first articles; and thus I came to be known as one of the few people in the world with an active interest in the Bactrian language.

This was no doubt the reason why, when the parchment illustrated here (fig. 1) came to light in 1991, the photos were forwarded to me. With a total of 28 almost complete lines on the two sides this was easily the most substantial text in cursive script which was known up to that time. I began to transliterate the text, following Gershevitch's system for the reading of the known letters and leaving gaps for the letters whose reading was still unknown. The meanings of a few common words were already known from the Bactrian coins and inscriptions; and some others could be tentatively interpreted on the basis of possible cognates in better-known Iranian languages. At some point it suddenly dawned on me that what I was reading was the beginning of a letter, using the same hyperbolic phrases with which I was familiar from Sogdian letters: "[To so-and-so] the lord, a thousand, ten thousand greetings and homage from so-and-so his servant. I have heard that your lordship is healthy, [therefore] I am [happy]"—and so on.

This first letter was already a revelation; but during the following years documents emerged from Pakistan or Afghanistan in a steady stream. Many were letters, some of them still sealed, with the text on the inside perfectly preserved. Others are economic documents, including tally sticks, or legal contracts. The latter are often preserved in two copies written on a single parchment, the upper copy being rolled up and sealed to avoid alteration and the lower copy left open to be read.

Many of these documents are dated, in an era which probably began in 223 AD, the inaugural year of the Sasanian dynasty of Iran. They range from the 4th century, in the period of Sasanian rule, to the late 8th century, well within the Islamic period, and cover all the centuries in between. Many of them also name the places where they were written, mainly in the principality of Rob, modern Rui in the Hindukush mountains, or in the cities of Guzgan, in north-west Afghanistan.

With this mass of new material, which has now grown to more than 150 items, it is no surprise that the remaining problems of reading the cursive script have simply disappeared. As Michael Ventris discovered in the case of Linear B, once you reach the stage where there is only one unidentified character in a word, it is comparatively easy to guess the value of that character. So I claim no particular credit for identifying the few letters which had not already been recognized by my predecessors. But of course, the decipherment of the script did not make the language instantly comprehensible. There was no bilingual, no Rosetta stone, and the texts still consisted almost entirely of unknown words, often in previously unknown grammatical forms, with no spaces to indicate where a new word begins. In other words, the decipherment of the script put scholars in the position in which Maricq found himself when the perfectly legible but incomprehensible inscription of Surkh Kotal came to light in 1957: the script could be read but the text could not yet be understood.

Of course, it is rather artificial to speak as if the decipherment of the script came first and the interpretation of the text came afterwards. In reality, the two processes proceeded hand in hand. As the reading of the letters became clearer so the meaning of the words emerged; and as the meaning emerged, so the readings could be improved.

I have spoken of meanings “emerging”, or even of a “revelation”, which no doubt sounds very unscientific. But in fact the way in which such a breakthrough is reached are the typical methods of all scientific enquiry: on the basis of context or a possible etymology, a hypothesis about the reading of a character or the meaning of a word is formulated, and then it must be tested, preferably in the light of new material. If the solution to a problem appears as a sudden flash of inspiration, this is merely because the confirmation sometimes follows the hypothesis so quickly. For example, a Bactrian letter always begins with one of two short words, or , the other of which appears a little further on within the first line or two. It does not take much imagination to guess that these must be the prepositions “to” and “from” and that sometimes the sender and sometimes the recipient is named first.<sup>1</sup> The first of these two words consists of letter-forms which had already been identified in the manuscripts from Turfan, and can be read immediately as *abo* “to”, a preposition known from the Surkh Kotal inscription. The other should therefore be the equally well-known *aso* “from” and its second letter, which had previously been read in various ways, should be a cursive form of *s* — a hypothesis easily checked by examining the many other words which contain the same character.

In the case just described the hypothesis, once formulated, was confirmed almost instantaneously. But of course things are not always so simple.

A problem which I grappled with for several years was the meaning of the word *masko*, which often appears near the end of the legal documents in a fixed phrase “then we shall pay the same fine as is written in/on (the) *masko*”. My first idea was to identify *masko* with the Old Persian word *maškā* “skin” (a word of Semitic origin), and to understand it as referring to the parchment on which the text is written.

This interpretation seemed plausible enough until the discovery in 1993 of a new Kushan inscription containing what is evidently an older form of the same word. In line 11 of this inscription I read the words: “he ordered images to be made of these gods who are written *maska*”. Since the inscription is written on stone, *maska* can hardly mean “parchment”. So I devised a new hypothesis, that is, a new translation “above”, supported by a new etymology (*m*- “the” + *-aska* = Sogdian *aska* “above”). The translation “images of these gods who(se names) are written (in) the above”, fits the context perfectly, since the list of the gods’ names immediately precedes the sentence I have quoted. This solution seems equally satisfactory in the contract with which we started, where the sentence quoted comes from the very end of a document and the amount of the fine is indeed mentioned “above”. But again a new discovery arrived to invalidate this second hypothesis. This was another parchment, a marriage contract.<sup>2</sup> The text begins by mentioning the date and the place of writing, followed by a reference to the witnesses “who witness the present document and (whose) signatures are written *masko*”. Here *masko* cannot mean “above” because the upper part of the document is perfectly preserved and contains no signatures. The only place where the signatures might be is at the bottom of the document, which is damaged but where one can indeed see traces of writing below the blank space where the seals were attached.

---

<sup>1</sup> It seems in fact that the sender only names himself first if his status is significantly higher than that of his addressee.

<sup>2</sup> Incidentally, this is the earliest dated Bactrian document (13 October 332?) and also one of the most remarkable: it records the marriage of a woman to two brothers at once, thus confirming later Chinese accounts of the practice of polyandry in Bactria.

So I devised yet another (I hope final) interpretation: “who witness the present document and (whose) signatures are written *hereupon*”. This reinterpretation doesn’t involve a change in the etymology, but only in the syntactic relationship between its elements: instead of understanding the initial *m-* as a definite article and the following *-aska/-asko* as equivalent to a noun, “the above”, one must take *m-* as a demonstrative “this” governed by *-aska/-asko* as a postposition “upon”, thus, “upon this, hereupon”.

As I mentioned, many of the legal documents exist in two copies, which often differ in small but interesting details. In one such case, the second (open) copy of the text contains our friend *masko* “hereupon” in the phrase: “as is written hereupon concerning the four boundaries”.<sup>3</sup> The parallel phrase in the first (sealed) copy contains a different expression: “as is written within (*bandaro*) concerning the four boundaries”. The choice of a different word, *bandaro*, which I interpret as “within”, from *b-* = *abo* “to, on, in” + *-andaro* = Middle Persian *andar* “inside”, may well be deliberate: in this case the details referred to are “inside” a scroll which is rolled up and sealed, while in the other they are “upon” the flat surface of the open copy.

In other instances we can determine the meaning of unknown words not by comparing two versions of the same text, but by comparing different, parallel texts. In Bactrian legal documents it is conventional to name the “houses” or “families” to which the parties to the contract belong. A typical expression is *kidomēno bono kadgo X razindo* “we whose estate (and) house they call X”. The vocabulary here includes *bono* “estate” (cf. Avestan *buna-*, Latin *fundus*), *kadgo* “house” (= Middle Persian and Parthian *kadag*) and *raz-* “to call, name”, a verb otherwise known only from Khotanese *rrāys-*. A later text replaces these words with synonyms: *kiddēno xano X girlindo* “you whose house they call X”. Here *xano* “house” is cognate with Sogdian *xānā*, Persian *xāna* etc., and *girl-* “to call, name” (with the typical Bactrian development of *l* from *\*d*) with Choresmian *rynd-*, Armenian *kard-*.

In order to interpret texts in a previously unknown language such as Bactrian the most basic requirement is an excellent knowledge of the cognate languages and their history, together with a broad familiarity with the cultural background of the area from which the texts derive and, of course, a good balance of ingenuity and common-sense. Through the application of these types of knowledge and skill to previously unreadable or incomprehensible texts, their meaning emerges, and with it a dead language comes back to life. In the case of Bactrian, we have reached the stage where the language is well enough understood to contribute to the study of the cognate languages, just as Mycenaean Greek, the language of the Linear B tablets, nowadays contributes to the understanding of the history of Classical Greek.

Despite the title of my talk, “From philology to history”, I am aware that I have in fact talked only about philology—about the process of deciphering and interpreting the Bactrian texts—not about what the historians can find in the texts once the philologists have done their work. To give even a sketch of what we can learn from the Bactrian documents and inscriptions about the political, economic, social and religious history of ancient Afghanistan would have required another hour at least; but I think you can imagine, even without my telling you, that the 200 or so documents and inscriptions which we can now read and, to a

---

<sup>3</sup>. The naming of the “four boundaries” of a property (i.e. east, west, north and south) is a feature which goes back via Aramaic contracts to ancient Mesopotamia.



large extent, understand inevitably provide a huge amount of information on every aspect of the history and culture of Afghanistan during the first millennium AD. We can follow the political history of Afghanistan over some eight centuries during which it was invaded many times; we learn of the practice of fraternal polyandry; we see that the traditional Zoroastrian religion faced competition from Buddhism. From the contracts we learn something of the legal system, with its roots in the Ancient Near East and the Hellenistic world; in the letters we have the first known references to the Afghan people. Some of these details are mentioned in external sources, such as the accounts left by Chinese Buddhist pilgrims, and some we could perhaps have guessed: but now we know them for sure, from the words which were put down in writing by those who actually lived in the region and which can now be read once again. It has been the task of the philologists to bring us to the point where the literal meaning of these words can be understood; now it is the turn of the historians to read between the lines and to bring us to a deeper understanding of the society in which the Bactrian texts were written.

Fig. 1. A Bactrian letter (xp = DOC. 1), Recto. Courtesy Professor D. N. Khalili.





# PLATE 8

Map: Afghanistan and adjacent regions, showing places mentioned in the Bactrian documents (  $\Delta$  ) and sites where Bactrian inscriptions have been found (  $\square$  ). Drawn by François Ory. © Nicholas Sims-Williams.

